Estimation of reference curves for fetal weight

Sandie Ferrigno[a], Myriam Maumy-Bertrand[b]
[a] Inria team BIGS Nancy - Grand-Est, Lab IECL, Université de Lorraine, France
[b] Lab IRMA, Université de Strasbourg, France

Objectives

Reference or standard curves are required in many biomedical problems. Values which lie outside the limits of these **reference curves** may indicated the presence of disorder.

Data are from the French **EDEN mother-child cohort** [1]. We are studying **fetal weight** (in grammes) that depends on the **gestional age** (GA) in the second and the third trimesters of mother's pregnancy.

Some classical parametric and semi-parametric methods as **polynomial** [2][3] and **LMS** method [4][5] or nonparametric methods as **kernel estimation** [6] are used to construct these curves. However, some of them requires strong assumptions.

EDEN Cohort [1]

EDEN is a mother-child cohort study investigating the prenatal and early postnatal determinants of child health and development.

2002 pregnant women were recruited before 24 weeks of amenorrhoea in two maternity clinics from middle-sized French cities (Nancy and Poitiers). From May 2003 to September 2006, 1899 newborns were then included.

Detailed information on phenotypes and exposures were collected through questionnaires from preg-nancy until 8 years and through clinical examinations of the mother and the child.

The main outcomes of interest are fetal (via ultrasound) and postnatal growth, adiposity development, respiratory health, atopy, behaviour and bone, cognitive and motor development.

The main early-life determinants studied are maternal nutrition, diet, eating behaviour, environmental pollutants and socioeconomic and psycho-emotional factors.

We acknowledge the commitment of the EDEN mother-child cohort study group and more particularly Barbara Heude which allowed us to use the data from the cohort.

Data and reference interval

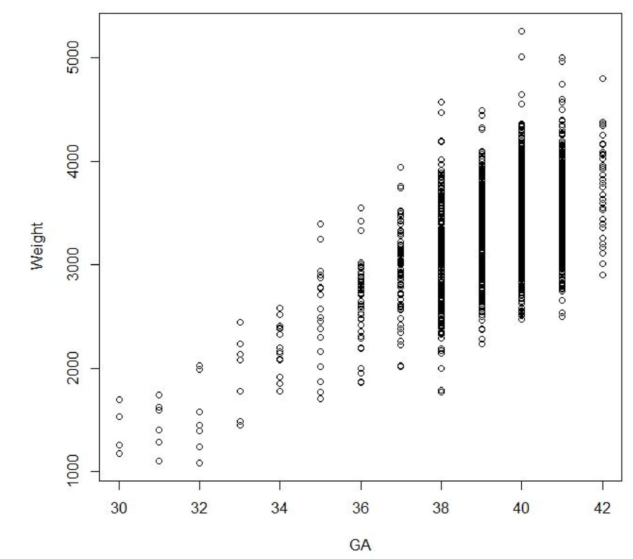

Figure 1: Scatter plot of fetal weight vs. GA.

An α% age-related reference interval is the range bounded by two centil curves which encompasses α% of the data at each age.

The estimate of the reference curves will be carried out from 30 GA because before we do not have enough data.

Polynomial method [2][3]

Polynomial regression is one of the most common parametric approach for modelling growth data espacially during the prenatal period. It is based on the asumption that at each t (GA) the measurement of weight has a normal distribution and that the mean and standard deviation vary smoothly with GA. The $C_\alpha(.)$ centile curves is estimated using

$$C_\alpha(t) = m(t) + Z_\alpha \sigma(t), \tag{1}$$

where Z_α is the normal equivalent deviate of size α corresponding to a particular centile and $m(.)$ and $\sigma(.)$ are the mean and standard deviation, respectively, at the required GA. In this study, these two functions are respectively estimated by polynomials of order three and one.

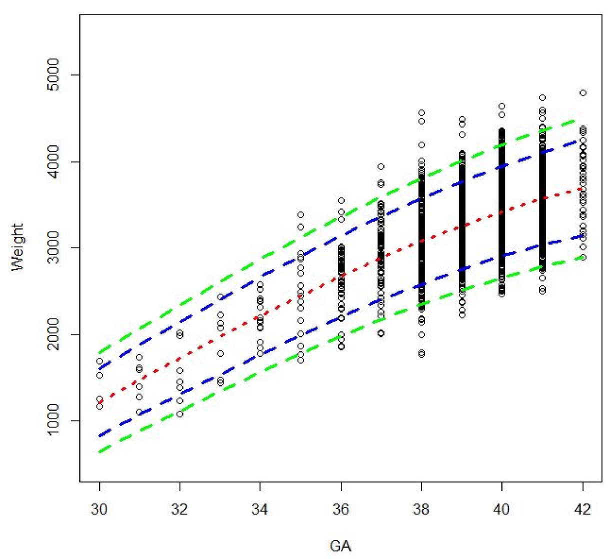

Figure 2: 3rd, 10th, 50th, 90th, 97th centile curves.

LMS method [4][5]

The LMS methods assumes that at each age the weight follows a normal distribution after a Box-Cox transformation. Therefore, data can be summarised by three GA-dependent functions that will be estimated and smoothed to obtain the location parameter curve $M(t)$, the coefficient of variation curve $S(t)$ and the value of the power needed to normalise the data at each age, $L(t)$. These three curves are fitted using cubic splines and by maximising the penalised likelihood. They together allow any centiles to be calculated

$$C_\alpha(t) = M(t)\left(1 + L(t) * S(t) * Z_\alpha\right)^{1/L(t)}, L(t) \neq 0$$

$$C_\alpha(t) = M(t)\exp(S(t) * Z_\alpha)), L(t) = 0$$

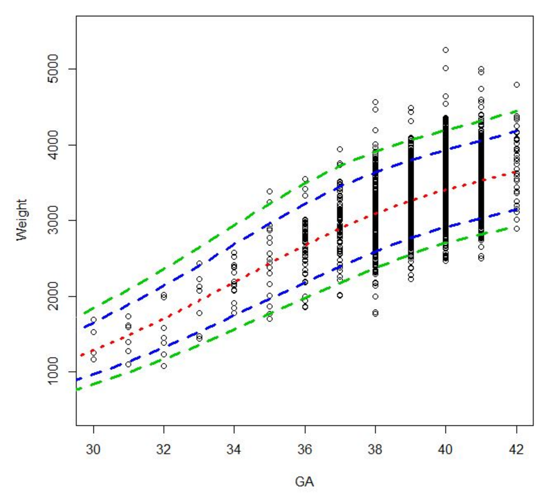

Figure 3: 3rd, 10th, 50th, 90th, 97th centile curves.

Nonparametric method [6]

Nonparametric method is used in equation (1) to estimate the GA-specific mean function $m(.)$ and the standard variance function $\sigma^2(.)$. These estimates are obtained using Nadaraya-Watson kernel estimation. $\hat{m}_n(.)$ and $\hat{\sigma}_n^2(.)$ are calculated from n realizations $\{(X_i, Y_i), i = 1, \ldots, n\}$ from the random variables X (GA) and Y (weight) by

$$\hat{m}_n(t) = \frac{\sum_{i=1}^n K\left(\frac{X_i - t}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{X_i - t}{h}\right)},$$

$$\hat{\sigma}_n^2(t) = \frac{\sum_{i=1}^n K\left(\frac{X_i - t}{h}\right) Y_i^2}{\sum_{i=1}^n K\left(\frac{X_i - t}{h}\right)} - \hat{m}_n^2(t),$$

where h and $K(.)$ are respectively a bandwidth parameter and a kernel function.

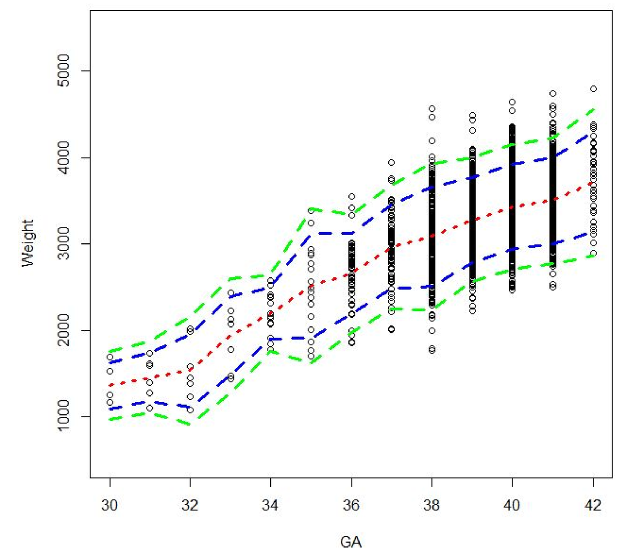

Figure 4: 3rd, 10th, 50th, 90th, 97th centile curves.

References

[1] B. Heude *et al.* Cohort profile: The eden mother-child cohort on the prenatal and early postnatal determinants of child health and development. *International Journal of Epidemiology*, pages 353–363, 2016.

[2] P. Royston and E. M. Wright. How to construct 'normal ranges' for fetal variables. *Ultrasound Obstet Gynecol*, pages 30–38, 1998.

[3] T. J. Cole. Fitting smoothed centile curves to reference data. *Journal of the Royal Statistical Society*, pages 385–418, 1988.

[4] E. O. Ohuma and D. G. Altman. Statistical methodology for constructing gestational age-related charts using cross-sectional and longitudinal data : The intergrowth-21 project as a case study. *Statistics in Medicine*, pages 1–20, 2018.

[5] T. J. Cole and P. J. Green. Smoothing reference centile curves : the lms method and penalized likelhood. *Journal of the Royal Statistical Society*, pages 1305–1319, 1992.

[6] A. Gannoun *et al.* References curves based on non-parametric quantile regression. *Statistics in Medicine*, pages 3119–3135, 2002.